Application of Statistical Estimation Procedures to the Identification Problem

by

R. P. Wishner and J. C. Lindenlaub

ABSTRACT

The method of maximum likelihood parameter estimation is applied to the problem of measuring parameters of an unknown linear filter or control system from input-output data when it is assumed that the output signal is corrupted with an additive Gaussian noise signal. The physical realization suggested by the integral formulation of the estimation technique is discussed and illustrated. Approximate expressions for the parameter estimates and the covariance matrix of the errors in the parameter estimates are obtained in the strong signal case. This analysis also has applications to the adaptive radar problem.

Introduction

The communication engineer has exploited the use of statistical parameter estimation techniques for a number of years, particularly, in problems concerning signal detection such as those that occur in radar. It has occurred to the authors that such techniques are useful to the control engineer as well. A natural application of statistical estimation techniques arises in adaptive control problems. The viewpoint that an adaptive process incorporates the ideas of system identification, decision, and modification[1] places this fact in evidence; the identification problem is nothing more than a problem in parameter estimation.

Some authors writing in the field of identification seem to have come close to noting this equivalence. Particular identification schemes employing crosscorrelation,[2] matched filters,[3] parameter tracking models,[4] etc. have been proposed, but we are unaware of any explicit statements to the effect that these specific realizations were motivated by the viewpoint of statistical parameter estimation. Once this equivalence between identification and statistical parameter estimation is noted a large amount of existing mathematical technique may be brought to bear on the identification problem. The purpose of this paper is, in effect, to transpose the method of maximum likelihood parameter estimation into the language of the control engineer. We consider the problem of estimating the unknown parameters of a control system when the signals are corrupted with noise.

---

1. Cooper, G. R., and Gibson, J. E., et al., "Survey of the Philosophy and State of the Art of Adaptive Systems," Technical Report No. 1, Contract AF33(616)-6890, PRF 2358, Purdue University, July, 1960.

2. Anderson, G. W., Aseltine, J. A., Marcini, A. R., and Sarture, C. W., "A Self-Adjusting System for Optimum Dynamic Performance," IRE National Convention Record, pt. 4, 1958.

3. Lichtenberger, W. W., "A Technique of Linear System Identification Using Correlation Filters," IRE Transactions - PGAC, Vol. AC-6, No. 2, May, 1961.

4. Margolis, M., and Leondes, C. T., "A Parameter Tracking Servo for Adaptive Control Systems," IRE Transactions - PGAC, Vol. AC-4, No. 2, November, 1959.

The basic identification problem is illustrated in Fig. 1. $K\,g(t; \underline{\alpha})$ is the impulse response of the unknown system; the constant K is a convenient scale factor, and the vector $\underline{\alpha}$ places in evidence the dependency of the impulse response upon the unknown parameters $\underline{\alpha} = \{\alpha_1, \alpha_2, \ldots \alpha_p\}$. It is assumed that $g(t)$ is realizable and that the observation time of the output, T, is choosen so that $g(t) \approx 0$ for $t > T$. The output signal $Ks(t; \underline{\alpha})$ is corrupted with an additive noise signal, $n(t)$, which is assumed to be Gaussian and have a known continuous autocorrelation function $R(t, s)$. Estimates of the unknown parameters K and $\underline{\alpha}$, are to be based upon measurements, of the output signal plus noise, $y(t)$, and the input signal, $x(t)$.

Maximum likelihood estimates of the set of unknown parameters, $\{\alpha_i\}$, are obtained. The set $\hat{\underline{\alpha}}$ which maximizes the conditional probability function $p(y|\underline{\alpha})$ is known as the maximum likelihood estimate of $\underline{\alpha}$. $p(y|\underline{\alpha})$ is considered to be a function of the $\{\alpha_i\}$, and as such, is called the likelihood function. The notation $L(\underline{\alpha})$ will be used to emphazize the dependence upon $\underline{\alpha}$. Maximum likelihood estimates have the advantage of yielding an efficient estimate, that is one with minimum variance, if such an estimate exists.[5] Excellent discussions of maximum likelihood as well as other parameter estimation techniques can be found in references (5), (6) and (7).

---

5. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, N. J., 1946.

6. Helstrom, C. W., Statistical Theory of Signal Detection, Pergamon Press, New York, N. Y., 1960.

7. Davenport, W. B., and Root, W. L., An Introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

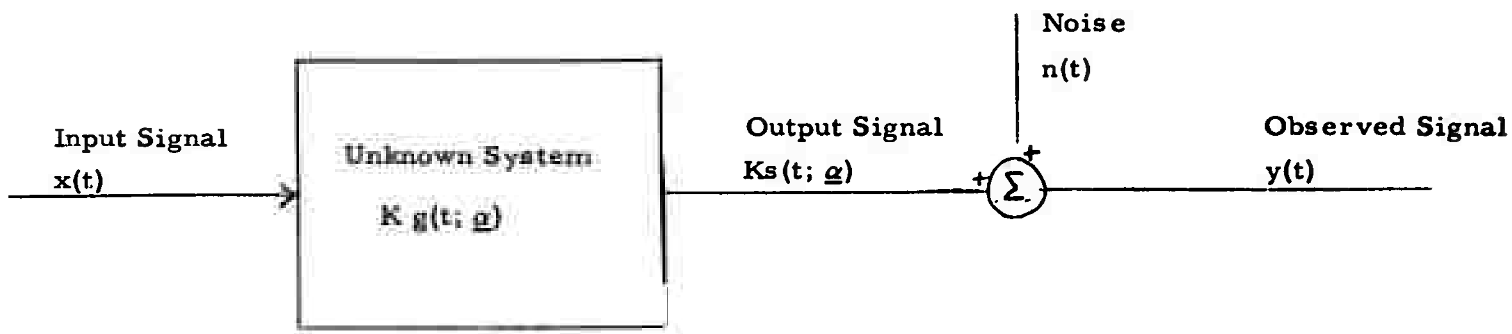

Fig. 1

The Identification Problem

Maximum Likelihood Estimates of α

Derivation of the maximum likelihood estimates can be obtained by expanding the noise process, as well as the other quantities of interest $x(t)$, $y(t)$, $s(t;\alpha)$ and $g(t;\alpha)$, in a series of orthogonal functions. Then $n(t)$, $x(t)$, etc. can be represented by the coefficients of this series. The Karhunen-Loeve expansion(7)(8)(9)(10) of $n(t)$ will be used because the coefficients of this series are uncorrelated. Thus, if $n(t)$ has a continuous correlation function $R(t, s)$, then $n(t)$ can be represented over the interval $(0, T)$ by the series

$$n(t) = \underset{N \to \infty}{\text{l.i.m.}} \sum_{k=1}^{N} n_k \phi_k(t) \tag{1}$$

Here the set of functions $\{\phi_k(t)\}$ (assumed to be a complete set) are the eigenfunctions associated with the eigenvalues, $\{\sigma_k^2\}$, of the integral equation

$$\int_0^T R(t, s)\, \phi_k(s)\, ds = \sigma_k^2\, \phi_k(t) \tag{2}$$

and the coefficients n_k (the observables of the noise process) are

$$n_k = \int_0^T n(t)\, \phi_k(t)\, dt \tag{3}$$

---

7. ibid.

8. Loeve, M., Probability Theory, D. Van Nostrand, Princeton, N. J., 1955.

9. Grenander, U., "Stocastic Processes and Statistical Inference," Arkiv For Matematik, 1950.

10. Kelly, E. J., Reed, I. S., and Root, W. L., "The Detection of Radar Echoes in Noise. I and II," J. Soc. Indust. Appl. Math., Vol. 8, No. 2, June, 1960, (I) and Vol. 8, No. 3, September, 1960, (II).

Hereafter, all expansions similar to Eq. 1 should be considered as convergent in the mean and the explicit notation used in Eq. 1 will be dropped. As mentioned above, the coefficients of the series in Eq. 1 have the properly

$$E(n_k\, n_l) = \sigma_k^2 \, \delta_{kl} \tag{4}$$

so that $n(t)$ may be represented by the sequence of uncorrelated random variables $\{n_k,\ k = 1, 2, \ldots\}$.

Expressions analogous to Eqs. 1 and 3 are assumed to exist for $x(t)$, $g(t;\underline{\alpha})$, $s(t,\underline{\alpha})$, and $y(t)$. The constant K is chosen so that

$$\sum_{k=1}^{\infty} \frac{s_k^2(\underline{\alpha})}{\sigma_k^2} = 1 \tag{5}$$

for all $\underline{\alpha}$. Note that this normalization effects the meaning of $s(t;\alpha)$ and $g(t;\alpha)$. The noise, $n(t)$, is assumed to be a zero mean Gaussian process. In this case the n_k's are Gaussian random variables with zero means and variances σ^2_k . Similarly the y_k's will be jointly Gaussian with means $K s_k$ and variances σ^2_k . The likelihood function for the Gaussian case is

$$L(\underline{\alpha}) = p(y\,|\,\underline{\alpha}) = \prod_{k=1}^{\infty} (2\pi\,\sigma_k)^{-1/2} \exp\left\{\frac{-\left(y_k - K s_k(\underline{\alpha})\right)^2}{2\sigma_k^2}\right\}$$

$$= \prod_{k=1}^{\infty} (2\pi\,\sigma_k)^{-1/2} \exp\left\{\frac{-\left(\cancel{y_k^2} - 2y_k K s_k(\underline{\alpha}) + K^2 s_k^2(\underline{\alpha})\right)}{2\sigma_k^2}\right\} \tag{6}$$

The $\underline{\alpha}$ which maximizes $L(\underline{\alpha})$ is desired. Since the first term in the exponential is not a function of the $\underline{\alpha}$ it may be disregarded. Also $L(\underline{\alpha})$ is maximized when $\ln L(\underline{\alpha})$ is a maximum so that the

desired set of estimates is the set that maximizes the expression

$$\sum_{k=1}^{\infty} \frac{\left(2y_k K s_k(\underline{\alpha}) - K^2 s_k^2(\underline{\alpha})\right)}{2\sigma_k^2}. \tag{7}$$

Completing the square in Eq. 7 and making use of Eq. 5 the estimation procedure can be written as

$$\max_{\underline{\alpha}} \max_{K} \left\{ \frac{1}{2} \left(\sum_{k=1}^{\infty} \frac{y_k s_k(\underline{\alpha})}{\sigma_k^2} \right)^2 - \frac{1}{2} \left(\sum_{k=1}^{\infty} \frac{y_k s_k(\underline{\alpha})}{\sigma_k^2} - K \right)^2 \right\} \tag{8}$$

The maximization over K can be done by inspection so that the estimate of K becomes

$$\hat{K} = \sum_{k=1}^{\infty} \frac{y_k s_k(\hat{\underline{\alpha}})}{\sigma_k^2} \tag{9}$$

The estimates $\hat{\underline{\alpha}}$ of $\underline{\alpha}$ are then chosen to satisfy

$$\max_{\underline{\alpha}} \left(\sum_{k=1}^{\infty} \frac{y_k s_k(\underline{\alpha})}{\sigma_k^2} \right)^2 = \max_{\underline{\alpha}} Y^2(\underline{\alpha}) \tag{10}$$

where $Y(\underline{\alpha})$ is defined implicitly.

The signal $K\,s(t;\underline{\alpha})$ is related to the unknown system impulse response and the input signal by the convolution integral

$$K\,s(t;\underline{\alpha}) = \int_{-\infty}^{\infty} x(\lambda)\,K\,g(t-\lambda;\underline{\alpha})\,d\lambda \tag{11}$$

and the coefficients $s_k(\underline{\alpha})$, can be expressed as

$$s_k(\underline{\alpha}) = \int_{-\infty}^{\infty} x(\lambda) \; g_k(\lambda;\underline{\alpha}) \, d\lambda \tag{12}$$

where

$$g_k \, (\lambda;\underline{\alpha}) = \int_{0}^{T} g(t - \lambda; \underline{\alpha}) \; \phi_k(t) \, dt \quad . \tag{13}$$

Note that $g_k(\lambda;\alpha)$ is zero outisde the interval $-T < \lambda < T$. Using these relations $Y(\underline{\alpha})$ of Eq. 10 can be written as

$$Y(\underline{\alpha}) = \sum_{k=1}^{\infty} \left\{ \sigma_k^{-2} \int_{0}^{T} y(t) \; \phi_k(t) \, dt \int_{-\infty}^{\infty} x(\lambda) \; g_k(\lambda;\underline{\alpha}) \, d\lambda \right\}$$

$$= \int_{0}^{T} dt \; y(t) \int_{-\infty}^{\infty} x(\lambda) \; g_1(t, \; \lambda;\underline{\alpha}) \, d\lambda \tag{14}$$

where $g_1(t, \; \lambda;\underline{\alpha})$ is defined as

$$g_1(t, \lambda;\underline{\alpha}) = \sum_{k=1}^{\infty} \frac{g_k(\lambda;\underline{\alpha}) \; \phi_k(t)}{\sigma_k^2} \quad , \qquad 0 \leqq t \leqq T \tag{15}$$

It can be shown, by direct substitution, that $g_1(t, \lambda;\underline{\alpha})$ satisfies the integral equation

$$\int_{0}^{T} R(t - s) \; g_1(s, \lambda;\underline{\alpha}) \, ds = g(t - \lambda; \; \underline{\alpha}) \quad , \quad 0 \leqq t \leqq T \tag{16}$$

Substituting Eq. 14 into Eq. 10 gives the following integral expression

for finding the maximum likelihood estimates of the $\{\alpha_i\}$.

$$\max_{\underline{\alpha}} \left\{ \int_0^T y(t) \int_{-\infty}^{\infty} x(\lambda)\ g_1(t, \lambda; \underline{\alpha})\ d\lambda\ dt \right\}^2 \tag{17}$$

Physical Interpretation

Let $\underline{\alpha}^o$, α_i^o and K^o denote the true value of the parameters $\underline{\alpha}$, α_i , and K . Maximum likelihood estimates of the α_i^o are obtained by multiplying the output signal (plus noise) of the unknown system by the output of a filter $g_1(t, \lambda; \underline{\alpha})$, and integrating this product over the observation time, T . Both the unknown system and g_1 are subjected to the same input signal. Because of the delay, or memory, of g and g_1 inputs between the times $-T$ and $+T$ affect the output signal during the interval $0, T$. (Recall that the observation time, T , was assumed to be longer than the significant duration of $g(t)$ so that inputs prior to $-T$ have a negligible affect upon the output during $0, T$.)

A physical realization of this process is illustrated in Fig. 2. The outputs from a bank of estimating filters, each with a different set of parameters $\{\alpha_i\}$, are integrated and squared. The set of parameters that corresponds to the channel with maximum output at time T is then the set of α_i which represents the maximum likelihood estimates of the true parameters, and the value of the signal at the output of the integrator is the estimate of K^o .

For the special case of white noise $R(t - s) = N_o\ \delta(t - s)$ so that the integral in Eq. (16) becomes trivial

$$g_1(t, \lambda; \underline{\alpha}) = \frac{1}{N_o}\ g(t - \lambda; \underline{\alpha}) \quad , \quad 0 \leqq t \leqq T \tag{18}$$

In this case $g_1(t, \lambda; \underline{\alpha})$ is a physically realizable time invariant filter. When $n(t)$ is not white the realization of g_1 is not as

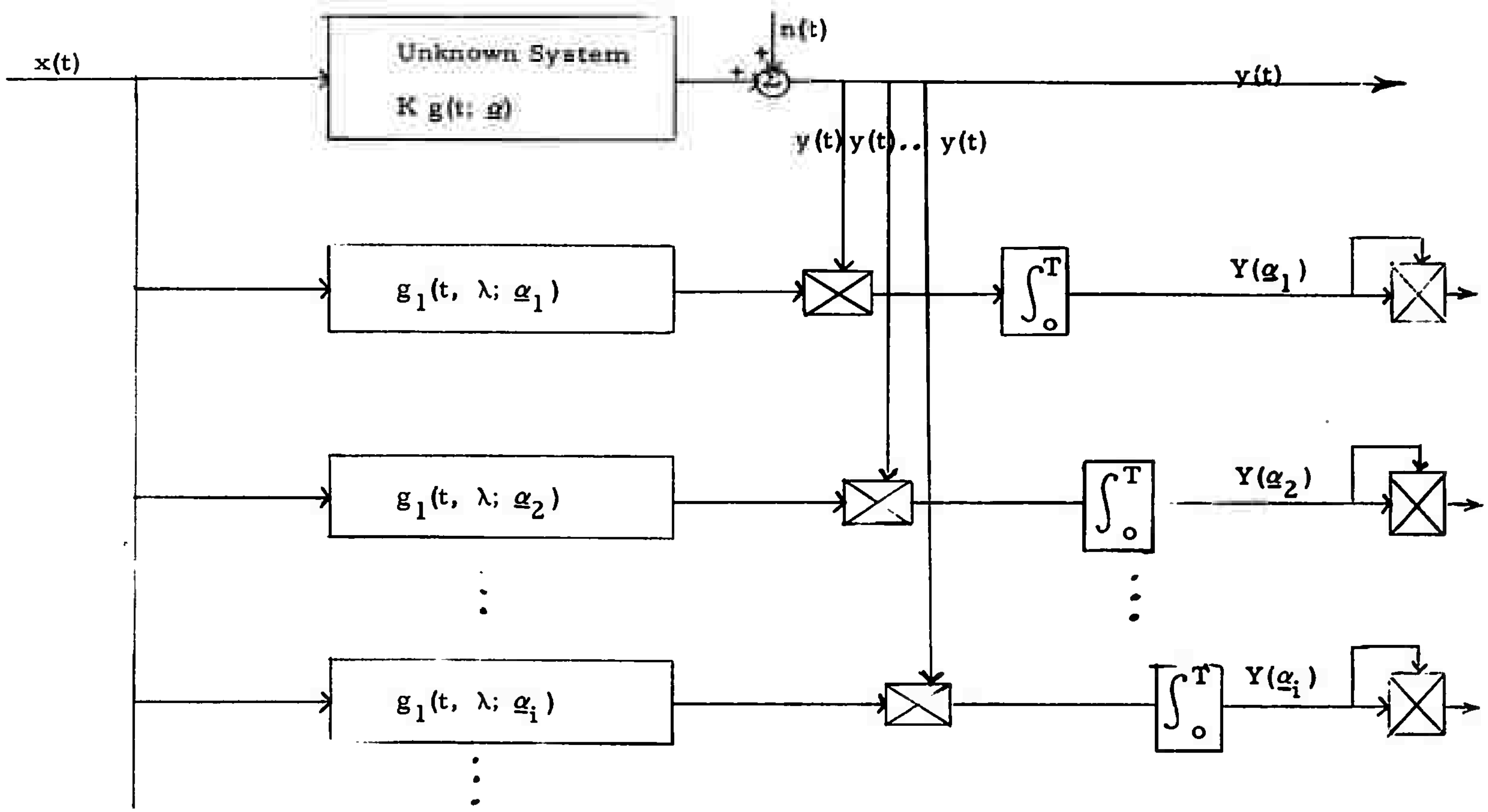

Fig. 2

Maximum Likelihood Estimator

straight forward. It is possible to approximate g_1 by a finite number of terms of the defining series, Eq. 15. The $\phi_k(t)$ and σ_k^2 are determined by the noise autocorrelation function $R(\tau)$, which is assumed to be known, and the $g_k(\lambda;\underline{\alpha})$ can be evaluated from the known form of $g(t)$ using Eq. 13. Such a realization is shown in Fig. 3.

The procedure for obtaining estimates of the α_i's for the non-white noise case is as follows. At $t = -T$ the time variable gians are "started", that is the gains are set equal to

$g_k \dfrac{(-T;\underline{\alpha})}{\sigma^2{}_k}$ and λ is allowed to traverse the interval $-T, +T$.

At $t = 0$ the integrators in Fig. 2 are reset, and finally at $t = T$ the outputs of each channel are examined and the channel with the maximum signal is determined. It is necessary to "start" the g_1 filter at $t = -T$ to insure that both g and g_1 have corresponding initial conditions at $t = 0$.

In general $g_1(t,\lambda)$ will not be realizable. However, Eq. 17 is equivalent to

$$\max_{\underline{\alpha}} \left\{ \int_T^{2T} y(t-T) \int_{-\infty}^{\infty} x(\lambda)\, g_1(t-T,\ \lambda;\underline{\alpha})\, d\lambda\, dt \right\}^2$$

It can be shown that $g_1(t-T,\ \lambda;\ \underline{\alpha})$ is always realizable. Thus, by delaying $y(t)$ by time T and replacing $g_1(t,\ \lambda;\ \underline{\alpha})$ by $g_1(t-T,\ \lambda,\ \underline{\alpha})$ a realizable system can be obtained.

Only the analytical form of the <u>product</u>, $K(\underline{\alpha})\, g(t;\underline{\alpha})$, has been assumed to be known; $g(t;\underline{\alpha})$ itself is not known and the multiplying constant depends upon both $\underline{\alpha}$ and $x(t)$ as well as the properties of the noise. For any given set of $\underline{\alpha}$ and $x(t)$ the constant K can be computed and $g(t;\underline{\alpha})$ found. At first sight this might seem discouraging because it suggest that the filters $g_1(t;\underline{\alpha})$ cannot be constructed until $x(t)$ is known. This difficulty can be avoided however by noting that it is possible to multiply both sides of

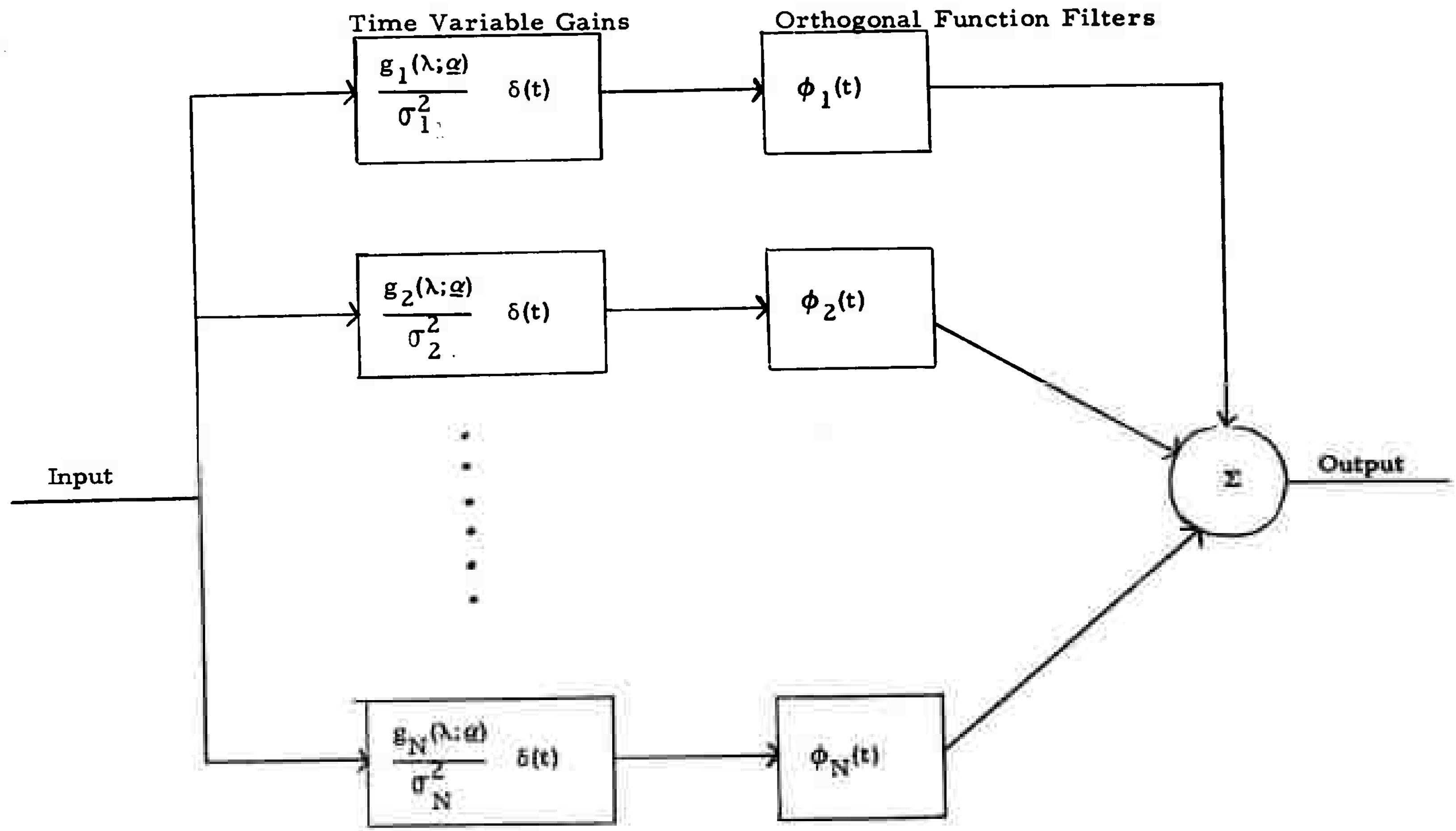

Fig. 3

Approximate Realization of $g_1(t, \lambda, \underline{\alpha})$

Eq. 16 by $c\,K(\underline{\alpha})$ and solve for $c\,K(\alpha)\,g_1(t,\lambda;\underline{\alpha})$ (c is an unknown constant *). Here it is not possible to separate $c\,K(\underline{\alpha})$ from $g_1(t,\lambda;\underline{\alpha})$ without a knowledge of the input, but the product $c\,K(\underline{\alpha})\,g_1(t,\lambda;\underline{\alpha})$ is independent of $x(t)$.

If a bank of filters $c\,K(\underline{\alpha})\,g_1(t,\lambda;\underline{\alpha})$ is used in Fig. 2 the outputs of each channel would be $c^2 K^2(\underline{\alpha})\,Y^2(\underline{\alpha})$ instead of $Y^2(\underline{\alpha})$. The desired outputs would be obtained by dividing each output by the appropriate $c^2 K^2(\underline{\alpha})$. Thus a scheme for computing $c\,K(\underline{\alpha})$ is required. Since the analytical from of $K(\underline{\alpha})\,g(t;\underline{\alpha})$ is known $c\,K(\underline{\alpha})$ can be computed for any $x(t)$ by observing $c\,K(\underline{\alpha})\,s(t;\underline{\alpha})$ (obtained from the output of an appropriate filter) and using Eq. 5. The maximum likelihood estimator then takes the form of Fig. 4. For the white noise case the "gain computer" is a rather simple device based on the relation

$$(c\,K)^2 \int_0^T s^2(t;\underline{\alpha})\,dt = (c\,K)^2 N_0$$

The left hand side of the above equation is the energy out of a filter with impulse response $c\,K(\underline{\alpha})\,g(t;\underline{\alpha})$ and can be easily obtained by a scheme such as shown in Fig. 5. For the non-white noise case a spectral analysis (with respect to the orthogonal set of functions $\phi_k(t)$) of $c\,K(\underline{\alpha})\,s(t;\underline{\alpha})$ is required so that the sum

$$c^2 \sum_k \frac{K^2(\underline{\alpha})\;s_k^2(t;\underline{\alpha})}{\sigma_k^2}$$

can be computed and $c\,K(\underline{\alpha})$ found so as to satisfy the normalization condition of Eq. 5. As mentioned above in the non-white noise case delays of time T may be necessary to insure realizability.

---

* The unknown constant c is introduced to emphasize the fact that only the form of $K\,g(t,\lambda;\alpha)$ is known, that is, it is known only to a scale factor, $1/c$.

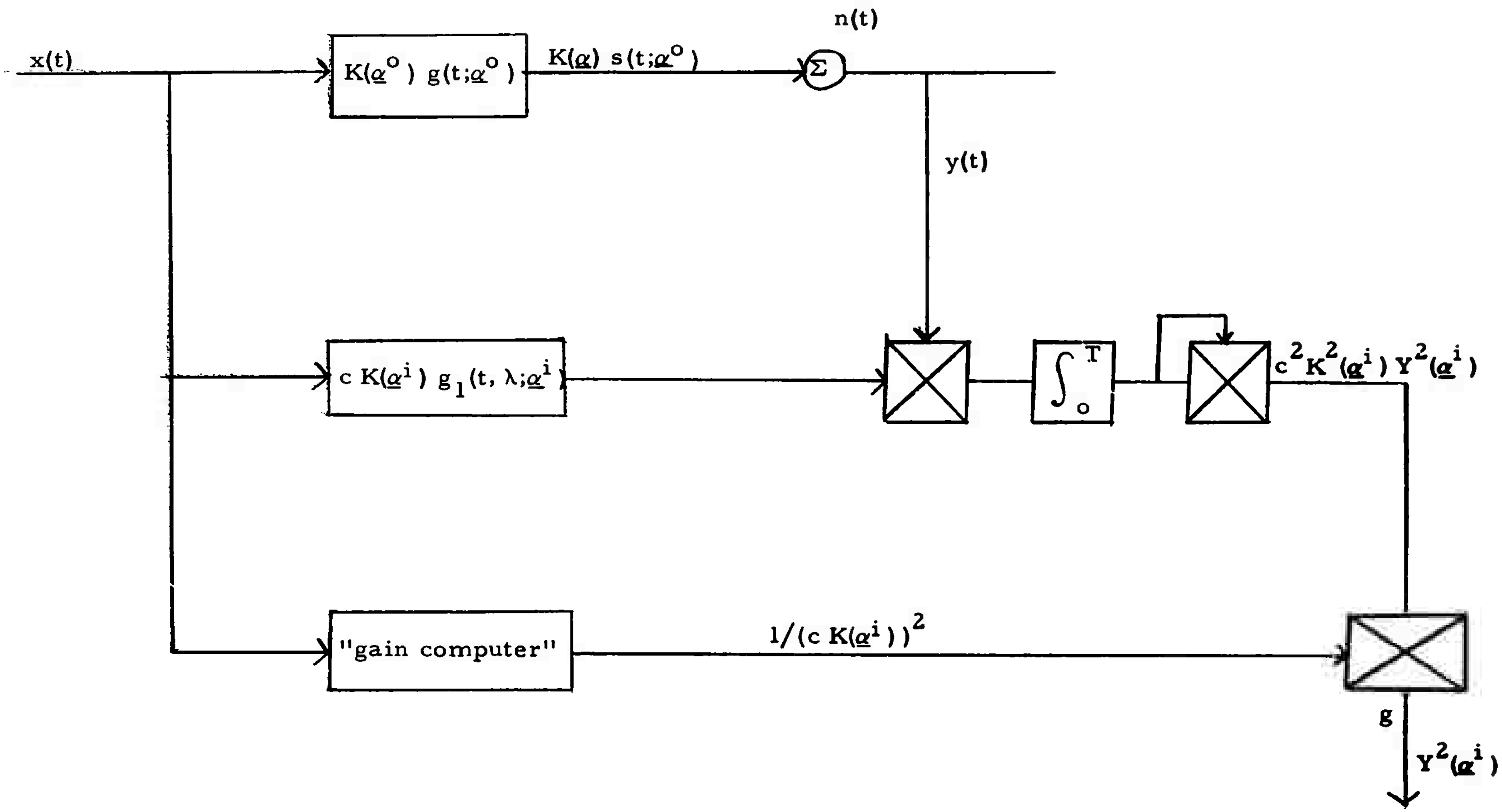

Fig. 4

i^{th} Channel of the Maximum Likelihood Estimator

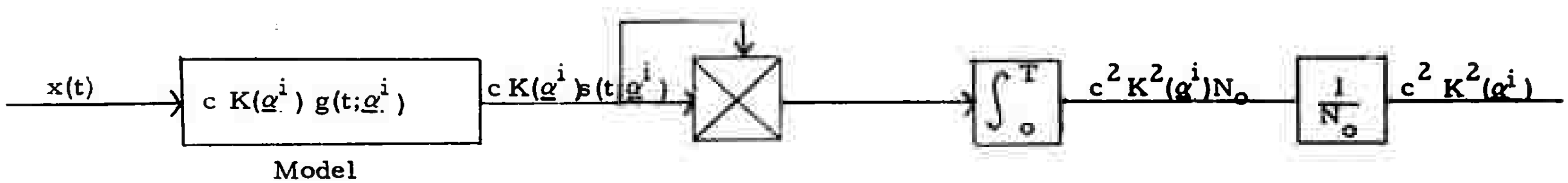

Fig. 5

Single Channel of "Gain Computer" for the White Noise Case

Estimation Errors

It is of interest at this point to consider the problem of finding an expression for the errors associated with the system identification. The errors will depend upon the observation interval T so that it will be possible to determine the observation time that is required for some specified error variance. This figure, the identification time, would be important, for instance, in considering the stability of an adaptive control loop.

Before becoming involved with the mathematical details, the general philosophy of the approach will be outlined. The "solution" for the set of values $\{\hat{\alpha}_i\}$ which maximizes Eq. 17 can be obtained by simultaneously setting the partial derivatives (c.f., Eq. 14)

$$\frac{\partial Y^2(\underline{\alpha})}{\partial \alpha_i} = 0 \qquad i = 1, 2, \ldots p$$

equal to zero and solving the resultant (in general nonlinear) set of equations in the $\{\hat{\alpha}_i\}$. This procedure is usually not practical. When the signal is sufficiently strong however $Y^2(\underline{\alpha})$ may be accurately represented, in the neighborhood of $\underline{\alpha}^o$, by the first few terms of a Taylors series expansion in the $\{\alpha_i\}$ variables, and approximations, $\{\tilde{\alpha}_i\}$, to the $\{\hat{\alpha}_i\}$ can be obtained by maximizing this expansion with respect to the $\{\alpha_i\}$. This is the procedure that is followed here. The approximate errors $\{\delta \tilde{\alpha}_i\} = \{\tilde{\alpha}_i - \alpha^o{}_i\}$ are then expressed to first order accuracy in n(t) and the variances $\sigma_{ij} = E(\delta \tilde{\alpha}_i \, \delta \tilde{\alpha}_j)$ are determined. The derivation presented here closely follows that of Kelly's[10] with specialization of the ambiguity function $Q(\underline{\alpha}, \underline{\alpha}^o)$ (defined below) to the control problem.

---

10. ibid.

Since

$$y(t) = K^o s(t; \underline{\alpha}^o) + n(t) \tag{19}$$

$Y(\underline{\alpha})$ (Eq. 14) can be expressed as

$$Y(\underline{\alpha}) = K^o \int_0^T s(t; \underline{\alpha}^o) \int_{-\infty}^{\infty} x(\lambda)\ g_1(t, \lambda; \underline{\alpha})\ d\lambda\, dt\ +$$

$$+ \int_0^T n(t) \int_{-\infty}^{\infty} x(\lambda)\ g_1(t, \lambda; \underline{\alpha})\ d\lambda\, dt \tag{20}$$

By defining

$$Q(\underline{\alpha}, \underline{\alpha}^o) = \int_0^T s(t; \underline{\alpha}^o) \int_{-\infty}^{\infty} x(\lambda)\ g_1(t, \lambda; \underline{\alpha})\ d\lambda\, dt$$

$$= \sum_{k=1}^{\infty} \sigma_k^{-2} \int_0^T s(t; \underline{\alpha}^o)\ \phi_k(t)\ dt \int_{-\infty}^{\infty} x(\lambda)\ g_k(\lambda; \underline{\alpha})\ d\lambda$$

$$= \sum_{k=1}^{\infty} \sigma_k^{-2}\ s_k(\underline{\alpha}^o)\ s_k(\underline{\alpha}) \tag{21}$$

and

$$N(\underline{\alpha}) = \int_0^T n(t) \int_{-\infty}^{\infty} x(\lambda)\ g_1(t, \lambda; \underline{\alpha})\ d\lambda\, dt$$

$$= \sum_{k=1}^{\infty} \sigma_k^{-2} \int_0^T n(t)\ \phi_k(t)\ dt \int_{-\infty}^{\infty} x(\lambda)\ g_k(\lambda; \underline{\alpha})\ d\lambda$$

$$= \sum_{k=1}^{\infty} \sigma_k^{-2} \, n_k \, s_k(\underline{\alpha}) \tag{22}$$

Equation 20 can be expressed as

$$Y(\underline{\alpha}) = K^o \, Q(\underline{\alpha}, \underline{\alpha}^o) + N(\underline{\alpha}) \tag{23}$$

Because of the normalization introduced in Eq. 5

$$Q(\underline{\alpha}, \underline{\alpha}) = \sum_{k=1}^{\infty} \frac{s_k^2(\underline{\alpha})}{\sigma_k^2} = 1 \tag{24}$$

and from the Schwarz inequality

$$Q(\underline{\alpha}, \underline{\alpha}^o) \leq 1 \tag{25}$$

Thus $Q(\underline{\alpha}, \underline{\alpha}^o)$ attains its maximum value at $\underline{\alpha} = \underline{\alpha}^o$. There may be other values of $\underline{\alpha}$ which maximize $Q(\underline{\alpha}, \underline{\alpha}^o)$. These values correspond to ambiguities in the parameter estimation problem. It is assumed here that either these ambiguities do not exist, or that they are resolved by other means.

Equation 24 guarantees that if the noise were zero $(y(t) = K\,s(t))$ for a particular observation $Y^2(\underline{\alpha})$ would be maximized by the set $\hat{\underline{\alpha}} = \underline{\alpha}^o$ and the gain estimate would be from Eqs. 9 and 10.

$$\hat{K} = Y(\hat{\underline{\alpha}}) = Y(\underline{\alpha}^o) = K^o \tag{26}$$

which is also the true value.

For the strong signal case the estimates will be close to the true values so that $Y^2(\underline{\alpha})$ can be expanded in a Taylor's series about $\underline{\alpha}^o$. Keeping terms only up to the quadratic term

$$Y^2(\underline{\alpha}) \approx Y^2(\underline{\alpha}^o) + \sum_{i=1}^{p} b_i\, \delta\alpha_i + \frac{1}{2} \sum_{i,j=1}^{p} c_{ij}\, \delta\alpha_i\, \delta\alpha_j \tag{27}$$

where

$$b_i = \left. \frac{\partial}{\partial \alpha_i} Y^2(\underline{\alpha}) \right|_{\underline{\alpha} = \underline{\alpha}^o} \tag{28}$$

and

$$c_{ij} = \left. \frac{\partial^2 Y^2(\underline{\alpha})}{\partial \alpha_i\, \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} \tag{29}$$

Setting the derivatives of Eq. 27 with respect to the $\{\alpha_i\}$ equal to zero and solving, one obtains for the deviation of the approximate estimate from the true parameter value

$$\tilde{\alpha}_i - \alpha_i^o = \delta\tilde{\alpha}_i = - \sum_{j=1}^{p} (C^{-1})_{ij}\, b_j \tag{30}$$

where $(C^{-1})_{ij}$ in the i, j element of C^{-1}, the inverse of $C = [c_{ij}]$. Substituting Eq. 30 back into Eq. 27

$$Y^2(\underline{\tilde{\alpha}}) = Y^2(\underline{\alpha}^o) - \frac{1}{2} \sum_{i,j=1}^{p} c_{ij}\, \delta\tilde{\alpha}_i\, \delta\tilde{\alpha}_j \tag{31}$$

The approximate error moments can be found by computing the $\delta\tilde{\alpha}_i$ (Eq. 30) to first order in the noise. Up to first order in noise (Eq. 23)

$$Y^2(\underline{\alpha}) = K^{o^2} Q^2(\underline{\alpha}, \underline{\alpha}^o) + 2K^o Q(\underline{\alpha}, \underline{\alpha}^o) N(\underline{\alpha}) \tag{32}$$

and

$$b_i = 2K^o \left. \frac{\partial N(\underline{\alpha})}{\partial \alpha_i} \right|_{\underline{\alpha} = \underline{\alpha}^o} \tag{33}$$

In obtaining Eq. 33 use was made of the fact that $Q(\underline{\alpha}^o, \underline{\alpha}^o) = 1$ and that because $Q(\underline{\alpha}, \underline{\alpha}^o)$ has a maximum at $\underline{\alpha} = \underline{\alpha}^o$

$$\left. \frac{\partial Q(\underline{\alpha}, \underline{\alpha}^o)}{\partial \alpha_i} \right|_{\underline{\alpha} = \underline{\alpha}^o} = 0$$

Since b_i has no zero order term, from Eq. 30 it can be seen that it will suffice to calculate only the zero order term of $(C^{-1})_{ij}$. To the required accuracy

$$c_{ij} = K^{o2} \left. \frac{\partial^2 Q^2(\underline{\alpha}, \underline{\alpha}^o)}{\partial \alpha_i \, \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} = 2K^{o2} \left. \frac{\partial^2 Q(\underline{\alpha}, \underline{\alpha}^o)}{\partial \alpha_i \, \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o}$$

$$= -2K^{o2} (M^{-1})_{ij} \tag{34}$$

where m_{ij} is defined as

$$m_{ij} = (M)_{ij} = \left(- \left[\left. \frac{\partial^2 Q(\underline{\alpha}, \underline{\alpha}^o)}{\partial \alpha_i \, \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} \right]^{-1} \right)_{ij} \tag{35}$$

Substituting Eqs. 33, 34, and 35 into Eq. 30 we obtain to the required accuracy

$$\tilde{\alpha}_i - \alpha^o{}_i = \delta \tilde{\alpha}_i = - \frac{1}{K^o} \sum_{j=1}^{p} m_{ij} \left. \frac{\partial}{\partial \alpha_j} N(\underline{\alpha}) \right|_{\underline{\alpha} = \underline{\alpha}^o} \tag{36}$$

Since $\hat{\alpha}_i - \alpha^o{}_i \approx \tilde{\alpha}_i - \alpha^o{}_i$ Eq. 36 yields an approximate expression for $\hat{\alpha}_i$ depending on the true value $\alpha^o{}_i$. The expected value of the

errors are

$$E(\delta\tilde{\alpha}_i) = -\frac{1}{K^o}\sum_{j=1}^{p} m_{ij}\, E\left(\frac{\partial N(\underline{\alpha})}{\partial \alpha_j}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\right) \tag{37}$$

$$= -\frac{1}{K^o}\sum_{j=1}^{p}\left\{m_{ij}\sum_{k=1}^{\infty}\sigma_k^{-2}\,\frac{s_k(\underline{\alpha})}{\alpha_j}\bigg|_{\underline{\alpha}=\underline{\alpha}^o} E(n_k)\right\} = 0$$

because $E(n_k) = 0$. Thus the $\delta\tilde{\alpha}_i$ are unbiased. The variances of the estimates are

$$E(\delta\tilde{\alpha}_i\,\delta\tilde{\alpha}_j) = \frac{1}{K^{o2}}\sum_{k=1}^{p}\sum_{\ell=1}^{p}\left\{m_{ik}\,m_{j\ell}\;\times\right.$$

$$\left.\times\, E\left(\frac{\partial N(\underline{\alpha})}{\partial \alpha_k}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\frac{\partial N(\underline{\alpha})}{\partial \alpha_\ell}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\right)\right\} \tag{38}$$

The expectation in Eq. 38 is

$$E\left(\frac{\partial N(\underline{\alpha})}{\partial \alpha_i}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\frac{\partial N(\underline{\alpha})}{\partial \alpha_j}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\right) =$$

$$= \sum_{k=1}^{\infty}\sigma_k^{-2}\,\frac{\partial s_k(\underline{\alpha})}{\partial \alpha_i}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\sum_{\ell=1}^{\infty}\sigma_\ell^{-2}\,\frac{\partial s_\ell(\underline{\alpha})}{\partial \alpha_j}\bigg|_{\underline{\alpha}=\underline{\alpha}^o} E(n_k\, n_\ell)$$

$$= \sum_{k=1}^{\infty}\sigma_k^{-2}\,\frac{\partial s_k(\underline{\alpha})}{\partial \alpha_i}\bigg|_{\underline{\alpha}=\underline{\alpha}^o}\frac{\partial s_k(\underline{\alpha})}{\partial \alpha_j}\bigg|_{\underline{\alpha}=\underline{\alpha}^o} \tag{39}$$

This last result may be further simplified by differentiating Eq. 24 twice giving

$$2 \sum_{k=1}^{\infty} \left\{ \sigma_k^{-2} \left[s_k(\alpha^o) \left. \frac{\partial^2 s_k(\underline{\alpha})}{\partial \alpha_i \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} + \right. \right.$$

$$\left. \left. + \left. \frac{\partial s_k(\underline{\alpha})}{\partial \alpha_i} \right|_{\underline{\alpha} = \underline{\alpha}^o} \left. \frac{\partial s_k(\underline{\alpha})}{\partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} \right] \right\} = 0 \qquad (40)$$

Using this result in Eq. 39

$$E\left(\left. \frac{\partial N(\underline{\alpha})}{\partial \alpha_i} \right|_{\underline{\alpha} = \underline{\alpha}^o} \left. \frac{\partial N(\underline{\alpha})}{\partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} \right) = - \sum_{k=1}^{\infty} \sigma_k^{-2} \, s_k(\underline{\alpha}^o) \left. \frac{\partial^2 s_k(\underline{\alpha})}{\partial \alpha_i \, \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o}$$

$$= - \left. \frac{\partial^2 Q(\underline{\alpha}^o, \underline{\alpha})}{\partial \alpha_i \quad \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o} \qquad (41)$$

Finally, using this last result in Eq. 38, results in a relatively simple expression for the covariances

$$E(\delta \alpha_i \; \delta \alpha_j) = - \frac{1}{K^{o2}} \sum_{k=1}^{p} \sum_{\ell=1}^{p} m_{ik} \, m_{j\ell} \left. \frac{\partial^2 Q(\underline{\alpha}^o, \underline{\alpha})}{\partial \alpha_i \; \partial \alpha_j} \right|_{\underline{\alpha} = \underline{\alpha}^o}$$

$$= \frac{1}{K^{o2}} \sum_{k=1}^{p} \sum_{\ell=1}^{p} m_{ik} \, m_{j\ell} \, (M^{-1})_{ij}$$

$$= \frac{1}{K^{o2}} \, m_{ij} \qquad (i, j = 1, \ldots p) \qquad (42)$$

Thus the approximate variances and covariances associated with measuring the parameters $\{\alpha_i\}$ depend only upon the second-partial derivative of $Q(\underline{\alpha}, \underline{\alpha}^o)$ evaluated at $\underline{\alpha} = \underline{\alpha}^o$. This result has a simple geometrical interpretation; the covariances depend only upon the

curvature of the likelihood function at its maximum. The variances depend upon the properties of the noise through the normalization contant K^o. When the noise is white $1/K^{o2} = N_o/E$, where E is the energy of the output signal.

Properties of the $Q(\alpha, \alpha^o)$ Function for the Control Problem

In the control problem the $Q(\alpha, \alpha^o)$ function (Eq. 21) is conveniently expressed as

$$Q(\underline{\alpha}, \underline{\alpha}^o) = \int_0^T \int_{-\infty}^{\infty} x(\lambda_1)\, g(t - \lambda_1; \underline{\alpha}^o)\, d\lambda_1 \int_{-\infty}^{\infty} x(\lambda_2)\, g_1(t, \lambda_2; \underline{\alpha})\, d\lambda_2\, dt$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\lambda_1)\, x(\lambda_2)\, q(\lambda_1, \lambda_2; \underline{\alpha}, \underline{\alpha}^o)\, d\lambda_1\, d\lambda_2 \qquad (43)$$

where

$$q(\lambda_1, \lambda_2; \underline{\alpha}, \underline{\alpha}^o) = \int_0^T g(t - \lambda_1; \underline{\alpha}^o)\, g_1(t, \lambda_2; \underline{\alpha})\, dt \qquad (44)$$

Note that $q(\lambda_1, \lambda_2; \underline{\alpha}, \underline{\alpha}^o)$ is zero when $|\lambda_1|$ or $|\lambda_2| > T$. The advantage of expressing $Q(\underline{\alpha}, \underline{\alpha}^o)$ in terms of $q(\lambda_1, \lambda_2; \underline{\alpha}, \underline{\alpha}^o)$ is that $q(\lambda_1, \lambda_2; \underline{\alpha}, \underline{\alpha}^o)$ is independent of the input signal and makes it possible to express the derivatives of Q in terms of the derivatives of q. Such a procedure simplifies the computations when it is desired to study the variances of the estimates for the different types of input signals occuring in control systems. If $x(t)$ should be a test signal used solely for identification purposes, then Eq. 43 would prove useful in a search for a test signal to minimize the variance of the error.

Conclusions

Maximum likelihood estimation techniques can be applied to the identification problem of control engineering, and a practical maximum likelihood estimator can be synthesized. For the case of white noise the realization is quite feasible. In the non-white noise case the realization is complicated by two factors, the solution of the appropriate integral equation leading to $g_1(t, \lambda; \alpha)$, and the realization of the estimating filter itself.

Analysis of the large signal case provides expressions for the variances of the maximum likelihood estimates. These calculations provide a means for evaluating the maximum likelihood estimation technique without the necessity of actually building or simulating the device. Also, since in the white noise case maximum likelihood estimation is equivalent to the classical method of least squares,[13] variances associated with a least square error type of identification technique.

As a further application of our analysis it should be noted that the control system problem as formulated herein is the same as the radar problem where the signal to be transmitted is not known a priori but only decided upon on the basis of past returns. Thus we have also treated the adaptive radar problem.

RPW:jd

---

13. Helstrom, C. W., Statistical Theory of Signal Detection, Pergamon Press, New York, N. Y., 1960.

UNCLASSIFIED

UNCLASSIFIED